

## Genetic Differentiation in South Amerindians Is Related to Environmental and Cultural Diversity: Evidence from the Y Chromosome

Eduardo Tarazona-Santos,<sup>1,3</sup> Denise R. Carvalho-Silva,<sup>1</sup> Davide Pettener,<sup>3</sup> Donata Luiselli,<sup>3</sup> Gian Franco De Stefano,<sup>4</sup> Cristina Martinez Labarga,<sup>4</sup> Olga Rickards,<sup>4</sup> Chris Tyler-Smith,<sup>5</sup> Sérgio D. J. Pena,<sup>1</sup> and Fabrício R. Santos<sup>2</sup>

Departamentos de <sup>1</sup>Bioquímica e Imunologia and <sup>2</sup>Biologia Geral, Universidade Federal de Minas Gerais, Minas Gerais, Brazil; <sup>3</sup>Area di Antropologia, Dipartimento di Biologia, Università di Bologna, Bologna; <sup>4</sup>Dipartimento di Biologia, Università di Roma "Tor Vergata," Rome; and <sup>5</sup>Department of Biochemistry, University of Oxford, Oxford, United Kingdom

The geographic structure of Y-chromosome variability has been analyzed in native populations of South America, through use of the high-frequency Native American haplogroup defined by the *DYS199-T* allele and six Y-chromosome-linked microsatellites (*DYS19*, *DYS389A*, *DYS389B*, *DYS390*, *DYS391*, and *DYS393*), analyzed in 236 individuals. The following pattern of within- and among-population variability emerges from the analysis of microsatellite data: (1) the Andean populations exhibit significantly higher levels of within-population variability than do the eastern populations of South America; (2) the spatial-autocorrelation analysis suggests a significant geographic structure of Y-chromosome genetic variability in South America, although a typical evolutionary pattern could not be categorically identified; and (3) genetic-distance analyses and the analysis of molecular variance suggest greater homogeneity between Andean populations than between non-Andean ones. On the basis of these results, we propose a model for the evolution of the male lineages of South Amerindians that involves differential patterns of genetic drift and gene flow. In the western part of the continent, which is associated with the Andean area, populations have relatively large effective sizes and gene-flow levels among them, which has created a trend toward homogenization of the gene pool. On the other hand, eastern populations—settled in the Amazonian region, the central Brazilian plateau, and the Chaco region—have exhibited higher rates of genetic drift and lower levels of gene flow, with a resulting trend toward genetic differentiation. This model is consistent with the linguistic and cultural diversity of South Amerindians, the environmental heterogeneity of the continent, and the available paleoecological data.

### Introduction

Our current state of knowledge about the tempo and mode of demographic events undergone by the ancestors of present-day Native Americans is full of uncertainties. During the 20th century, these issues have been sources of vigorous debate between anthropologists and have also involved population geneticists (Salzano and Callegari-Jaqués 1988; Crawford 1998; Powell and Neves 1999). In recent years, most molecular-genetic studies of Amerindians have focused on the clarification of aspects of their Asian origins, by comparing the genetic structure of the two continental groups. These studies have used female- and male-specific DNA-lineage markers (i.e., mtDNA and Y chromosome, respectively) to ascertain

the genetic structure of the two groups, to infer the population size of the first colonists (Schurr et al. 1990; Ward et al. 1991; Pena et al. 1995), and to infer the number or geographic sources of the discrete migration waves that peopled the Western Hemisphere (Torroni et al. 1993; Merriwether et al. 1996; Karafet et al. 1999; Santos et al. 1999b). Moreover, they have attempted to date the most recent common ancestors of Asian-Amerindian molecular lineages (Torroni et al. 1993; Forster et al. 1996) and the migration waves by dating of the Native American-specific lineages (Underhill et al. 1996) or of the population expansion following the colonization of the new continent (Bonatto and Salzano 1997). However, these inferences concerning the aforementioned genetic or demographic events are often not consistent among themselves or with available linguistic, archeological, and paleoanthropological data (Crawford 1998; Powell and Neves 1999).

Studies of Y-chromosome variability in Native American populations have been helpful in disentangling some aspects of the genetic history of these populations. They have shown the existence of a major northern-Asian founder haplotype, originally called "II-A" (Pena et al. 1995; Santos et al. 1996), the carriers of which

Received February 7, 2001; accepted for publication April 9, 2001; electronically published May 15, 2001.

Address for correspondence and reprints: Dr. Fabrício R. Santos, Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Avenue Antônio Carlos 6627, CP 486, 31.270-010, Belo Horizonte, Minas Gerais, Brazil. E-mail: fsantos@icb.ufmg.br

© 2001 by The American Society of Human Genetics. All rights reserved. 0002-9297/2001/6806-0020\$02.00

likely migrated from central Siberia (Karafet et al. 1999; Santos et al. 1999b). Moreover, a Native American-specific mutation, which is derived with respect to the II-A haplotype, has also been reported (Underhill et al. 1996). This derived allele, *DYS199-T*, has been found at very high frequencies in almost all the Native American populations surveyed for Y-chromosome variation (Bianchi et al. 1998).

In contrast to these intensive efforts to clarify the earliest origin of Native Americans, molecular-genetic studies at more-local geographic scales, such as those examining North, Central, or South America, are scanty (Kolman and Bermingham 1997; Mesa et al. 2000). Although some Native American populations have been studied for Y-chromosome microsatellites that are able to differentiate among them (Bianchi et al. 1998; Ruiz-Linares et al. 1999), the geographic structure of genetic variation has not been used to infer aspects of population history. For this reason, we decided to examine the genetic structure of Y-chromosome variability in South America and to interpret our results through a consideration of geographic information.

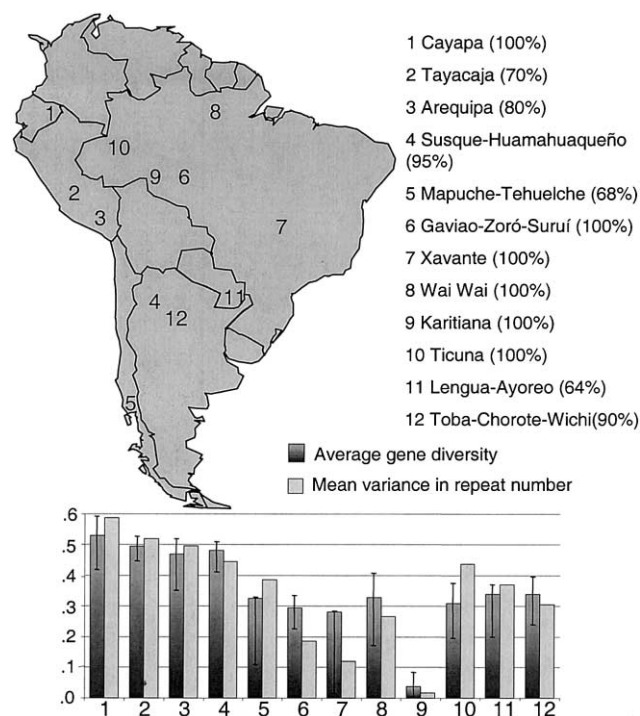
In the present study, we focus our analysis on South Amerindian populations, increasing its resolution by studying several Y-chromosome markers and analyzing a geographically broad sample of native individuals, in a search for geographic patterns of genetic structure at two hierarchical levels: within populations and among populations. We tested two simple null hypotheses: (i) that there are no differences, in within-population variability, among the different populations and (ii) that patterns of variation match the isolation-by-distance (IBD) model (Wright 1943). On the basis of our results, we propose a qualitative model for the evolution of the male lineages of South Amerindians. Finally, the inferred evolutionary factors (i.e., the combined effects of genetic drift and gene flow) that would have shaped the genetic structure of these populations are correlated with historical and paleoecological events. A contrasting picture emerges from the analysis of Y-chromosome variability in South America, where populations have interacted with a variety of biomes during a period of  $\geq 12,000$  years, yielding very differentiated social structures, including those of the Andean empires, the Amazonian tribes, and the societies of the central Brazilian plateau.

## Subjects, Material, and Methods

### Populations, Samples, and Markers

We studied 192 South Amerindian individuals (see below). From this sample we analyzed, in detail, 169 individuals carrying the Native American-specific C→T transition in *DYS199*, tested by the protocol described by Santos et al. (1999a). This mutation defines Y-chro-

mosome haplogroup 18 (Santos and Tyler-Smith 1996), to which the analyses are restricted in the present study. Figure 1 shows the frequency of this mutation in the following populations: Cayapa ( $n = 26$ ), from the tropical forests of Ecuador; Tayacaja ( $n = 44$ ) and Arequipa ( $n = 15$ ), from the Peruvian Andes; Ticuna ( $n = 32$ ), Wai Wai ( $n = 5$ ), Gavião ( $n = 17$ ), Zoró ( $n = 4$ ), Suruí ( $n = 13$ ), and Karitiana ( $n = 8$ ), from Brazilian Amazonia; and Xavante ( $n = 5$ ), from the central Brazilian plateau. Data for aggregated Brazilian native populations were published by Carvalho-Silva et al. (1999), whereas data for the Peruvian and the Cayapa populations are unpublished. The Cayapa and Tayacaja samples have been described elsewhere (Rickards et al. 1994; Luiselli et al. 2000). The Arequipa sample originates from the rural highlands of the homonymous department (i.e., the political units into which Peru is divided), where native Quechua populations are predominant. All individuals sampled were informed about our objectives and consented to the anonymous use of samples for non-profit research. Our statistical analyses also include the following populations studied by Bianchi et al. (1998): Susque and Huamahuqueño ( $n = 28$ ), from the Andes of northern Argentina (also see the report by Dipierri et al. [1998]); Toba, Chorote, and Wichi ( $n = 21$ ), from



**Figure 1** Geographic distribution of samples, frequencies of haplogroup 18, and statistics (gene diversities with their 95% CI and variance in number of repeats) of within-population variability for haplogroup 18, averaged over six Y-chromosome-linked microsatellite loci (*DYS19*, *DYS389A*, *DYS389B*, *DYS390*, *DYS391*, and *DYS393*). Some samples are pooled, when geographically and linguistically close.

northern Argentina; Mapuche ( $n = 3$ ) and Tehuelche ( $n = 5$ ), from central Chile; and Lengua ( $n = 5$ ) and Ayoreo ( $n = 5$ ), from southern Paraguay. For our statistical analysis, we pooled populations represented by small samples ( $n \leq 5$ ) that were close to one another geographically and linguistically. Figure 1 shows the geographic localization of these populations and reports the aggregations performed.

In principle, it is not correct to infer aspects of the history of an entire population on the basis of variability of one or a few haplogroups that account for only a partial component of overall genetic variability. However, in this specific case, given the very high frequency of haplogroup 18 in South Amerindians (the haplogroup was present in 88% of our sample; see reports by Underhill et al. [1996] and Bianchi et al. [1998] for other frequencies), the male history of Native American populations is highly correlated with the history of this haplogroup. Hence, by excluding the remaining rare haplogroups from the analysis, we lose very little information and gain accuracy. The current state of knowledge about the worldwide distribution of haplogroups other than 18 does not always allow for discrimination between aboriginal haplogroups and those that arrived after the European conquest. Since our aim is to make inferences about pre-Columbian history, the exclusion of haplogroups other than 18 prevents confounding effects due to recent gene flow from Europe or Africa.

The variability of Y chromosomes belonging to haplogroup 18 was analyzed through use of the following six Y-chromosome-linked microsatellites: *DYS19*, *DYS389A*, *DYS389B*, *DYS390*, *DYS391*, and *DYS393*. They were typed in two PCR multiplex reactions, as suggested by Carvalho-Silva et al. (1999) and Santos et al. (1999a). The alleles were resolved in an ALF (Pharmacia) automatic fluorescent DNA sequencer, and their size was established by the fragment-manager software AlleleLinks (version 1.2, Pharmacia) and scored as suggested by Kayser et al. (1997).

#### Analysis of Data

The genetic structure of Y-chromosome variability in South America was analyzed at two hierarchical levels—within populations and among populations—by statistical techniques based on the infinite-allele and stepwise-mutation models and nonparametric analysis. To use geographic information in our analysis, geographic coordinates were entered, if available, or were inferred for each sample. Although errors of approximation could be introduced by this procedure, these are negligible for our analysis at the continental level.

#### Within-Population Variability

We used two measures of diversity to test the null hypothesis of no geographic differences in within-pop-

ulation variability: (1) average gene diversity among the six microsatellite loci (formula 8.6 in the report by Nei [1987]), and (2) the mean variance in the number of repeats among them. In the first case, significant differences among populations were tested, using 95% confidence intervals (CIs) constructed by the bootstrap technique (sampling the haplotypes 5,000 times), using the software GENETIX (Belkhir et al. 1998). In the second case, we tested the significance of differences among populations or groups of populations by performing a nonparametric ANOVA for independent samples on the distribution of individual average deviances among the six loci, calculated, for any chromosome, as  $(\frac{1}{6}) \sum_l (k_l - \mu_{kl})^2$ , where  $l$  corresponds to the locus considered, which, in this case, can vary from 1 to 6;  $k_l$  is the number of repeats for locus  $l$  in the chromosome; and  $\mu_{kl}$  is the mean number of repeats for locus  $l$  in the population to which the individual belongs. The reader can verify that the mean variance in repeat numbers in a population can be calculated either as the average of locus-specific variances or as the mean of the individual average deviances, calculated using the above formula.

#### Among-Population Variability

We applied three approaches to identify geographic structure of genetic variability among populations: genetic distance, the analysis of molecular variance (AMOVA), and spatial-autocorrelation analyses.

Two types of genetic distance were calculated. The first is simply the  $F_{st}$  (Weir and Cockerham 1984), which is the among-population component of genetic variance and is based on the infinite-allele model. The second is the  $R_{st}$ , proposed by Slatkin (1995) as a molecular version of  $F_{st}$  under the stepwise-mutation model. Moreover, Carvalho-Silva et al. (1999) have shown that, within haplogroup 18, locus *DYS19* has a significantly lower mutation rate than the do other five loci analyzed. This means that, on average, the expected coalescence time between two chromosomes is longer if they differ by a given number of repeats for the *DYS19* locus than it is for two chromosomes that differ to a similar degree at any of the other five loci analyzed here. To exclude the possibility that this heterogeneity in mutation rates could bias our results, we also calculated—as a way of double checking—a weighted version of  $R_{st}$ , hereafter called “ $R_{stw}$ ,” giving a weight of 2 to locus *DYS19* and a weight of 1 to the other loci. We also confirmed trends observed in the genetic-distance analysis by means of AMOVA (Excoffier et al. 1992), a nonparametric method that allows an assessment of the partitioning of genetic variance at different hierarchical levels. AMOVA was performed by the software Arlequin 2.0 (Schneider et al. 2000).

Genetic-distance matrices were graphically summarized by nonmetric multidimensional scaling (NM-MDS) (Kruskal 1964), through use of the software STATISTI-

CA. NM-MDS uses an iterative process to transform a similarity/dissimilarity matrix into distances represented in an Euclidean  $n$ -dimensional space. This method was preferred because it does not require that the data be distributed in a multivariate normal fashion or that the relationships be linear (James and Culloch 1990).

To test for possible correlation between genetic and geographic distances, pairwise correlation coefficients were calculated between the genetic distances ( $F_{st}$ ,  $R_{st}$ , and  $R_{stw}$ , as a double-checking measure) and geographic distances, the latter being calculated, at the Great Circle Distances Web site, as the great circle distance, expressed in km, from the geographic coordinates of populations. The significance of correlations was calculated by random distribution of samples (1,000 times) among the locations considered. This was done through use of the Mantel test (Mantel 1967), by the procedure implemented in the software GENETIX (version 3.0).

The IBD model (Wright 1943) was also considered, as a more specific null hypothesis of geographic structure of Y-chromosome variability in South America. To test the IBD model, we used a spatial-autocorrelation analysis, specially developed for molecular data, called "AIDA" (autocorrelation index for DNA analysis; Bertorelle and Barbujani 1995). It calculates a normalized similarity index, called "II," at different geographic-distance classes. The plot of II indices versus geographic-distance classes is called a "correlogram," the shape of which can be related to specific evolutionary scenarios (Sokal 1979). A decreasing correlogram is expected under the IBD model, resulting from positive, significant II values at short distances changing to nonsignificant ones at large distances (Barbujani 1987). The significance of II values was assessed by a randomization test, as described by Bertorelle and Barbujani (1995). It should be pointed out that AIDA assayed the geographic structure of genetic variability on the basis of between-individual comparisons, rather than between-population ones. Hence, AIDA could be a more powerful tool, for detection of geographic patterns of genetic variability, than is correlation between genetic and geographic distances. AIDA was performed, by the homonymous software (available from G. Bertorelle's Web site), through the conversion of Y-chromosome microsatellite haplotypes to arrays of binary digits whose number of pairwise differences is equal to the interhaplotype differences in number of repeats. In this case, we also double checked our results by performing a weighted AIDA, based on the same criteria used for  $R_{stw}$  distances.

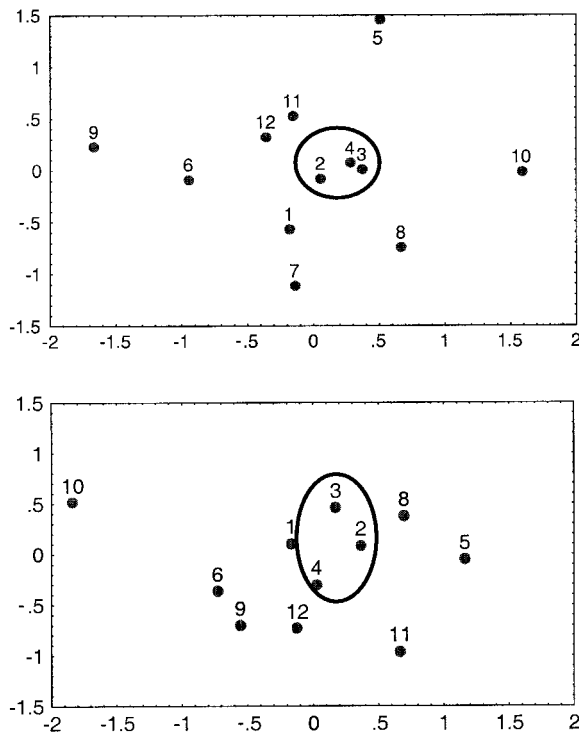
## Results

The microsatellite haplotypes found in the individuals analyzed, as well as their frequencies in the eight aggregated populations, are presented in the Appendix. Hap-

lotype 13.17.10.24.10.13, composed of the modal alleles in Native Americans (Bianchi et al. 1998; Ruiz-Linares et al. 1999), is the most widespread and the second-most frequent in the sample. Haplotype 13.16.10.25.10.15 exhibits the highest absolute frequency but seems restricted to the Ticuna population.

### *Patterns of Within-Population Variability*

Figure 1 presents the statistics used to test within-population variability for the 12 populations considered. A broad comparison of within-population variability indices between populations shows that the Cayapas (population 1 in fig. 1) of the extreme western portion of Amazonia (near the Ecuadorian Andes) and the Andean populations (Tayacaja, Arequipa, and Susque-Huamahuqueño; populations 2–4 in fig. 1), all found in the western part of the continent, have higher gene diversities and variances in number of repeats than do the other populations. This is true even when it is taken into consideration that eastern populations show wider CIs in average gene diversities, most likely related to their small sample sizes. This finding suggests the hypothesis of relatively high population variability in the Andean region. To further test this assertion, we performed two additional tests. First, we agglomerated the samples into two groups: Andean (populations 2–4 in fig. 1) and eastern (populations 6–12 in fig. 1), and calculated the mean gene diversity and (to be more conservative) their 99% CIs. The resultant average gene diversities were .489 (CI .448–.525) and .389 (CI .343–.425), for Andean and eastern groups, respectively (i.e., they showed non-overlapping confidence intervals). Although Chakraborty et al. (1988) suggested that gene diversity is relatively robust with respect to artificial agglomeration of populations, we should discuss the possibility that the difference found between Andean and eastern groups may be a statistical artifact of agglomeration. Despite the fact that both groups of populations have similar sample sizes (92 vs. 113), the eastern group includes more populations, which are also genetically more differentiated than the Andean ones (see results below, regarding among-population variability). This fact would create a stronger artificial Wahlund effect in the eastern group, inflating its gene diversity. This means that, if a bias exists in association with the agglomeration procedure, it acts against our expectation, thereby making our result more robust. Second, in order to use molecular information, we compared the above Andean and eastern groups by assessing the distributions of the individual mean deviances. We rejected the null hypothesis of no differences between the means of both distributions, in favor of the alternative hypothesis of higher deviance mean values for Andean populations (Kruskal-Wallis ANOVA, one-tail test,  $H = 29.16$ ,  $P < .05$ ). Moreover,



**Figure 2** Bidimensional representation of the genetic-distance matrices between South Amerindian populations, obtained by NM-MDS. *Top*,  $F_{st}$  genetic distance. *Bottom*,  $R_{st}$ . Populations are denoted by the numbers reported in figure 1, and the Andean populations are enclosed within the ellipse.

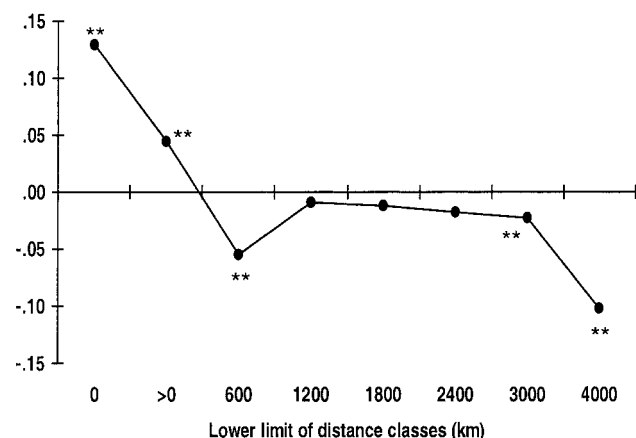
the Cayapa population, settled in the tropical forest very close to the Andes, also shows a high level of variability. Altogether, these results strongly suggest that populations inhabiting the western part of South America exhibit high levels of within-population variability compared to those found in eastern populations.

*Patterns of Among-Population Variability*

Figure 2 displays the two-dimensional configuration of the NM-MDS for the  $F_{st}$  and  $R_{st}$  distances. The genetic-distance matrices are available online at the Web site of the Laboratório de Biodiversidade e Evolução Molecular (at the institution with which the corresponding author is affiliated) or can be sent by the authors upon request. Since pairwise genetic distances based on completely linked loci could be associated with high evolutionary variances, we only considered general features of the genetic-distance distributions. Irrespective of the genetic-distance matrix used ( $F_{st}$ ,  $R_{st}$ , or  $R_{st}$ ), the most striking result is the homogeneity of the Andean populations (Tayacaja, Arequipa, and Susque-Huamahuaqueño; populations 2–4 in fig. 2), which exhibit very low and nonsignificant ( $P > .05$ ) genetic distances. In contrast, the non-Andean populations are scattered in the NM-MDS graph, con-

sistent with large genetic distances between them and between Andean and non-Andean populations. These results, based on pairwise genetic-distance analysis, are strongly confirmed by AMOVA. When Andean populations are considered as a group, the among-population component of molecular variance ( $\Phi_{st}$ ) does not achieve statistical significance ( $\Phi_{st} = .024$ ,  $P = .107$ ; when the Telhueche sample is included in the group,  $\Phi_{st} = .035$ ,  $P = .076$ ). Conversely, when eastern populations are pooled, the percentage of genetic variance corresponding to among-population comparisons is astonishingly higher ( $\Phi_{st} = .312$ ,  $P < .01$ ).

The correlation between genetic ( $F_{st}$ ,  $R_{st}$ ) and geographic distances does not evidence any significant association for the Y-chromosome variability in the considered sample ( $r = -.08$  with  $P = .58$ , for  $F_{st}$ ;  $r = .06$  with  $P = .29$ , for  $R_{st}$ ). This does not change when the  $R_{st}$  distance is used. Conversely, the autocorrelation analysis seems to be more informative. Figure 3 shows the correlogram obtained from the analysis of 236 South Amerindian Y chromosomes. Alternative correlograms constructed using a reasonable number of different distance classes, and either containing the same number of comparisons in any distance class or weighting the locus information as a function of locus-specific mutation rates, give the same shape, the same pattern of significance, and very similar values of autocorrelation indexes. The following characteristics emerge from the spatial autocorrelation analysis: (1) there is a general and significant resemblance between Y chromosomes at lower distance classes (0–600 km); (2) there is a depression in the correlogram at the third distance class,



**Figure 3** Spatial correlogram describing the geographic pattern of the microsatellite variability associated with haplogroup 18 in South American native populations. The X-axis represents geographic distance classes; the Y-axis represents  $I$  values. Two asterisks (\*\*) indicate that  $P < .01$ , and an absence of asterisks indicates that values are not significantly different from 0.

which contains comparisons between the highly differentiated Ticuna and five other populations; (3) despite the higher values of autocorrelation obtained for shorter distances, the correlogram shows a depression and significant negative values for the latter distance classes, which are not predicted by a pure IBD model (Sokal 1979; Barbujani 1987)—however, a categorical rejection of the IBD model would require the study of more samples that are widely distributed across the continent; and (4) the negative and significant values of autocorrelation observed in the latter distance classes would suggest the existence of a north-to-south cline, which (we speculate) could be related to the expansion of the first colonists of the continent. However, these negative and significant  $F_{ST}$  values could be strongly dependent on the small Chilean samples considered (Telhueche and Mapuche). Again, given the paucity of available samples, it is not possible to discriminate between a sample-specific characteristic and a continental expression of the structure of genetic variability.

## Discussion

In the present study, we have tested some simple null hypotheses posited to identify patterns of geographic structure of Y-chromosome variability in South American native populations. The most striking pattern we found is that the Andean and Cayapa populations show significantly higher levels of within-population and lower levels of among-population variability when compared with populations from Brazil, Paraguay, and northeastern Argentina. The results of genetic-distance and autocorrelation analyses seem to lead to the rejection of a pure isolation-by-distance model, and more-complex regional patterns of drift and gene flow should be invoked to explain most of the Y-chromosome genetic variability in South America.

The number of chromosomes analyzed in this study ( $n = 236$ ) represents the most widely distributed South Amerindian sample that has been considered so far. From a logistic point of view, South Amerindian Y chromosomes are difficult to sample, because urban centers have shown a predominance of European Y chromosomes (Carvajal-Carmona et al. 2000; Carvalho-Silva et al. 2001) and because high frequencies of native Y chromosomes seem to be restricted to rural populations, which are often geographically isolated. Although data about Colombian native populations have been published recently (Ruiz-Linares et al. 1999), they have been presented in an aggregated way, which prevents their use in analyses of geographic structure of genetic variability. Therefore, considering the sample size, we have used several cautionary procedures to render solid our conclusions. First, we double checked results about within-population variability, by pooling our samples

and controlling the effects of artificial agglomeration of populations. Second, we double checked results regarding among-population variability, through use of genetic distance analysis, AMOVA, and AIDA. In contrast to classical genetic-distance analysis, AMOVA and AIDA use molecular information, assessing the number of differences between haplotypes and increasing the number of comparisons considerably. Finally, we based our conclusions on results for which statistical significance has been accurately tested, and on general patterns of genetic variability rather than on specific population results.

## *Inferences about Evolutionary Factors*

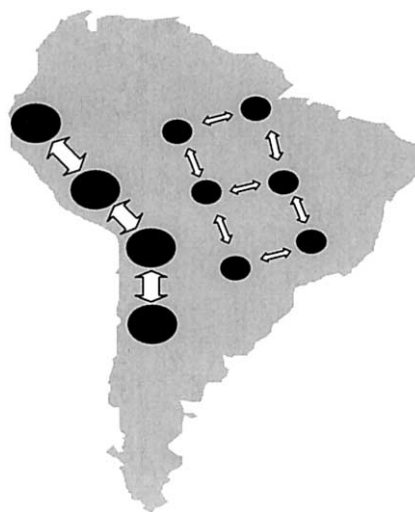
The higher variability within the Andean and Cayapa populations, which inhabit the western part of South America, suggests relatively greater effective population sizes. Three nonexclusive scenarios could be considered in order to explain these differences: (1) the ancestors of the western populations could have arrived in more ancient times than did the other populations under consideration; (2) the western populations could have been larger than the other populations (in terms of long-term effective population size) and (3) the western populations could have exchanged more genes among themselves than the other populations (i.e., the neighborhood-size model). Archaeological data do not support the first scenario, at least for the Andean populations, because the most ancient sites in the Andean region are not older than Brazilian sites (Sandweiss et al. 1998; Dillehay 1999). Moreover, an extensive settlement of the Andean region, which would not have been possible when glaciers were present at low altitudes, must have begun after the end of last glaciation (Bonavia 1991). Conversely, a combination of scenarios 2 and 3 seems more plausible for the Andean populations. Agriculture in South America was widespread in the Andean region before it spread in the eastern part of the continent (Harlan 1971; Bonavia 1991), and Andean populations built complex societies that acquired the greatest level of socioeconomic development in pre-Columbian South America. These historical facts are expected to have led to larger populations (Cavalli-Sforza et al. 1994), and when the Spaniards arrived in the 16th century, the Andean area and neighboring coasts indeed had the highest demographic densities in South America (scenario 2; see reviews by Sánchez Albornoz [1992] and Crawford [1998]). Furthermore, the fact that the shortest genetic distances correspond to comparisons between Andean populations is in accordance with the postulated higher levels of gene flow among them (scenario 3). On the other hand, the above explanations cannot account for the highest diversity being found in the Cayapa population, which inhabits western Amazonia close to the Andes and is differentiated genetically and culturally from

Andean populations (see fig. 2). Anthropological studies suggest an Amazonian origin of the Cayapas (Rickards et al. 1994). Although the Cayapa population exhibits mitochondrial DNA variability levels comparable to those of other Amazonian groups (Rickards et al. 1999), this is not true for Y-chromosome microsatellite haplotype variability. This can be explained by some level of Andean male contribution to the Cayapa gene pool, which is in agreement with ethnographic reports about Andean influences on the ethnogenesis of the Cayapas (Barrett 1925). In accordance with these multiple sources of male gene flow (Andean and Amazonian), the Cayapas would be expected to exhibit high population diversity.

#### *Evolutionary Model of South Amerindian Y-Chromosome Variability*

Our results concerning the structured genetic variation of Y chromosomes in South America can lead to a qualitative model to explain the evolution of male lineages in this region, on the basis of contrasting patterns of genetic drift and gene flow. This model is depicted in figure 4. In the western part of the continent, associated with the Andean region, populations exhibit larger effective sizes and higher gene-flow levels between them, which implies a trend towards the homogenization of the gene pool. In contrast, eastern populations, settled in the Amazonian region, the central Brazilian plateau, and the Chaco region, exhibit higher rates of genetic drift and lower levels of gene flow, with a resulting trend toward genetic differentiation. This model implies that South Amerindian populations should be considered as two groups evolving at different rates. Since the population-genetics methods that are used to make quantitative inferences about population history or population structure usually assume equal effective population sizes among populations, our findings suggest that the application of these kinds of analyses to the entire continent would not be appropriate.

In the present study, we have identified different patterns of drift and gene flow acting on western and eastern South Amerindian populations. These findings are consistent with results obtained from the multivariate analysis of genetic structure in South America, performed by Luiselli et al. (2000) and Simoni et al. (2000) through use of classical markers. They found that the central Andes represent a wide area that is free of barriers to gene flow, whereas they found zones of sharp genetic discontinuities in the eastern part of South America and weak but significant barriers to gene flow between the western and eastern parts of the continent. However, we did not detect a significant differentiation in haplogroup 18 microsatellite variability among the two groups formed by the Andean and eastern populations (among-group component of genetic variation  $\Phi_{ct} = .018, P = .288$ , estimated



**Figure 4** Diagram of evolutionary forces shaping the genetic structure of Y-chromosome variability in South America. Circle sizes roughly indicate the relative effective sizes of the represented populations. Arrow sizes denote gene-flow levels.

by AMOVA). This means, in the wide sense, that the same Y-chromosome alleles are present at similar frequencies in the Andes and the eastern part of South America, but that they are differentially distributed within those areas. This partial discordance could be related to the different samples used in both studies, the sex-specific correlation of Y-chromosome data, or the different evolutionary mechanism underlying the evolution of classical markers and haplotypes of Y-chromosome microsatellites. Notwithstanding these partial differences, both studies coincide in suggesting a differential east-west spatial organization of the gene pool of South Amerindian people. On the other hand, data regarding the highly variable Cayapa population suggest that some “suture populations” can exist, receiving genes from both the west and the east. Further studies will be necessary to verify their existence.

#### *Historical and Paleoecological Correlations*

The high level of genetic similarity among Andean populations indicates a strong correlation between genetic variability and environmental-cultural diversity. The Tayacaja, Arequipa, and Susque-Huamahuaqueño populations are settled in the Andean region, a set of mountains and valleys that exhibit a wide latitudinal and altitudinal distribution. From an ecological point of view, the Andean area is homogeneous, in the sense that, at a given altitude, latitudinal displacements find similar environments. This geographic area has been involved in a unique cultural process for the past 12,000 years, which has led to a relative cultural and linguistic homogeneity compared with eastern native populations of

South America, settled in the Amazonian region, the central Brazilian plateau, and the Chaco area (Sánchez-Albornoz 1992). This cultural resemblance among Andean populations could have favored intensive gene exchanges among them, even considering that they are not very close geographically. This is relevant because, in general, the high-altitude environment is associated with genetic differentiation and isolation (Cavalli-Sforza et al. 1994).

Our model is also compatible with paleoecological data. The Pleistocene-Holocene transition (10,000–12,000 years ago), at the end of the last glacial era, produced a great environmental impact and has been the most important paleoecological event of the past 20,000 years in South America (Ab'Sáber 1990). During the last glacial era, tropical forests were restricted to a few refugia, whereas the remaining lands, currently occupied by Amazonia, constituted a savanna environment. In this situation, the levels of gene flow between eastern populations could have been greater than they are now. When the Holocene began, the tropical forests that had been limited to the refugia expanded (Ab'Sáber 1990; Bonavia 1991), probably producing a barrier to intensive gene flow, which may subsequently have been limited to fluvial basins. It could have generated, at least partially, the observed trend toward genetic differentiation between Amazonian populations.

For a thorough interpretation of our results, we should discuss the possibility that the observed patterns of genetic differentiation may have been shaped after the European conquest (i.e., during the past 5 centuries). We believe that a number of considerations make this possibility unlikely. First, during the 3 centuries following the arrival of Europeans, indigenous populations of South America underwent a severe bottleneck (Sánchez-Albornoz 1992), and it is possible to argue that it has been more severe in the eastern part of the continent compared to in the Andean area, yielding the contrasting within-population genetic diversities found in the present study. However, even if a bottleneck produces an almost immediate effect on parameters closely related to the number of rare alleles, it has been shown that reduction of gene diversity (i.e., in average expected heterozygosity) began several generations later (Maruyama and Fuerst 1985; Cornuet and Luikart 1996). Therefore, the recent demographic depletion undergone by Amerindian populations 20–25 generations ago could not account for the differences in gene diversity evidenced in the present study, which are more likely to be related to more-ancient (i.e., pre-Columbian) demographic events.

On the other hand, during the centuries following the arrival of the European conquerors, there were dislocations of native populations in the Andean area. Could these displacements (which occurred during the past 20–25 generations) have determined the population

structure currently observed in the Andes? The recent and intensive displacements of natives in the Andean region have been constituted by short- and long-range migrations with a clear and predominant direction: people leave the rural areas to go to medium and large cities. Moreover, there is extensive gene flow of native people between large cities (Sánchez-Albornoz 1992). This means that if studies of the genetic structure of Andean populations are based on samples collected in medium or large cities, the results could be seriously biased toward homogenization of the Andean gene pool. However, this is not the case in the present study. In fact, our Peruvian samples (Tayacaja and Arequipa) have been collected from individuals belonging to farming communities, which settled in the highlands of the Andes before A.D. 1500 (also see reports by Luiselli et al. [2000] and Pettener et al. [1998], for genetic and historical information regarding the Tayacaja sample). The other two Andean samples from Argentina—Susque and Huamahuqueño—belong to small semiurban centers that are likely to have received gene flow from the surrounding farming communities (Dipierri et al. 1998). Therefore, our study is based on populations that constitute the source, rather than the destination, of these recent migratory movements. Furthermore, in addition to being separated by large geographic distances, Peruvian and Argentinian populations have belonged to different political units during the past 500 years (after conquest, they have been part of the Peruvian and Rio de La Plata viceroyalties, respectively). These considerations are also valid with respect to forced displacements, related to mineral-mining activities, that Andean populations underwent after conquest: mining centers founded during the colonial period could have received genetic contributions from near and far locations, but, again, this is not the case for samples collected for the present investigation. Hence, it is unlikely that the genetic resemblance observed between Peruvian and Argentinian samples could be due to gene-flow events that occurred during the 20–25 generations that have elapsed since the conquest. This resemblance is likely to be correlated with the unique linguistic, cultural, and historical evolutionary process that has been taking place in this extensive region during the past 10,000 years.

In the present study, we have made historical inferences on the basis of population-genetic analyses. In principle, a phylogeographic approach would be more powerful for the inference of population histories, often allowing for discrimination between scenarios based on genetic drift and gene flow (Templeton 1998), a distinction that is not possible in the present analysis. This approach could be used if an acceptable molecular phylogeny could be reconstructed without a lot of ambiguities. This is not the case for the six Y-chromosome microsatellites used in this study of South Amerindians. Recently, Forster et al.





**Table A1 (continued)**

HAPLOTYPE <sup>a</sup>	FREQUENCY IN POPULATION <sup>b</sup>									TOTAL
	1	2	3	6	7	8	9	10		
13.16.11.24.10.13		1								1
13.16.11.24.10.14	1									1
13.16.11.24.11.13	4									4
13.16.11.25.10.13	2									2
13.16.12.24.11.13	1									1
13.16.09.23.10.13	1									1
13.16.10.25.10.14								1		1
13.17.09.23.11.13		1								1
13.17.09.24.10.12						2				2
13.17.09.24.11.13		1								1
13.17.10.23.10.13		2		3						5
13.17.10.24.10.13		2	2			1	7			12
13.17.10.24.10.14		2								2
13.17.10.24.11.13	2						1			3
13.17.10.24.11.14		1								1
13.17.10.24.10.13				9						9
13.17.10.24.11.13				1						1
13.17.10.25.10.13				5						5
13.17.10.25.10.14			1							1
13.17.10.26.10.13		1								1
13.17.11.21.10.13		1								1
13.17.11.23.10.14		5	3							8
13.17.11.24.10.13	1	1	1							3
13.17.11.24.11.13				5						5
13.18.09.23.10.12						1				1
13.18.09.24.10.13	1	1								2
13.18.09.25.10.13	2									2
13.18.09.25.11.13	1									1
13.18.10.23.10.13								1		1
13.18.10.24.10.13		2						1		3
13.18.10.24.10.14	1	1								2
13.18.10.24.11.13				2						2
13.18.10.25.10.13	1	1								2
13.18.11.20.10.13		1								1
13.18.11.21.10.13		1								1
13.18.11.24.09.13				1						1
13.18.11.24.10.13	1							1		2
13.18.11.24.11.13	1			1						2
13.18.11.24.11.14	1									1
13.18.11.25.10.13				1						1
13.19.10.24.10.13				1						1
13.19.11.24.10.13								5		5
14.15.10.24.10.13		2								2
14.16.09.24.11.13		1								1
14.16.10.24.11.13		1								1
14.16.11.24.10.13		1								1
14.16.11.25.10.13			1							1
14.17.10.23.11.13		1								1
14.17.10.24.10.13		1								1
14.17.10.24.10.14		1								1
14.17.10.24.11.13										1
Total sample size	26	44	15	34	5	5	8	32	1	169

<sup>a</sup> The order of loci in each haplotype is *DYS19*, *DYS389A*, *DYS389B*, *DYS390*, *DYS391* and *DYS393*.

<sup>b</sup> Populations are denoted by the numbers reported in figure 1.

## Electronic-Database Information

URLs for data in this article are as follows:

Giorgio Bertorelle's Home Page, <http://www.unife.it/genetica/Giorgio/giorgio.html> (for AIDA software)

Great Circle Distance, [http://members.tripod.com/paul\\_kirby/appletgreatcircle/greatc.html](http://members.tripod.com/paul_kirby/appletgreatcircle/greatc.html) (for calculation of great circle distances)

Laboratório de Biodiversidade e Evolução Molecular, <http://www.icb.ufmg.br/lbem/Y-data> (for genetic distance matrices)

## References

- Ab'Sáber A (1990) Paleoclimas quaternários e pre-história da América tropical I. *Revista Brasileira de Biologia* 50:805–820
- Barbujani G (1987) Autocorrelation of gene frequencies under isolation by distance. *Genetics* 117:777–782
- Barrett S (1925) The Cayapa Indians of Ecuador. Indian notes and monographs. Vol 40. Heye Foundation, New York
- Belkhir K, Borsa P, Goudet J, Chicky L, Bonhomme F (1998) GENETIX, logiciels pour Windows pour la génétique des populations. Laboratoire Génome et Populations, Université de Montpellier II, Montpellier
- Bertorelle G, Barbujani G (1995) Analysis of DNA diversity by spatial autocorrelation. *Genetics* 140:811–819
- Bianchi NO, Catanesi CI, Bailliet G, Martinez-Marignac VL, Bravi CM, Vidal-Rioja LB, Herrera RJ, López-Camelo JS (1998) Characterization of ancestral and derived Y-chromosome haplotypes of New World native populations. *Am J Hum Genet* 63:1862–1871
- Bonato SL, Salzano FM (1997) A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. *Proc Natl Acad Sci USA* 94:1866–1871
- Bonavia D (1991) Perú: Hombre e Historia: De los orígenes al siglo XV. Vol I. EDUBANCO, Lima
- Carvajal-Carmona LG, Soto ID, Pineda N, Ortíz-Barrientos D, Duque C, Ospina-Duque J, McCarthy M, Montoya P, Alvarez VM, Bedoya G, Ruiz-Linares A (2000) Strong Amerind/white sex bias and a possible sephardic contribution among the founders of a population in northwest colombia. *Am J Hum Genet* 67:1287–1295
- Carvalho-Silva DR, Santos FR, Hutz MH, Salzano FM, Pena SD (1999) Divergent human Y-chromosome microsatellite evolution rates. *J Mol Evol* 49:204–214
- Carvalho-Silva DR, Santos FR, Rocha J, Pena SD (2001) The phylogeography of Brazilian Y-chromosome lineages. *Am J Hum Genet* 68:281–286
- Cavalli-Sforza L, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton
- Chakraborty R, Smouse PE, Neel JV (1988) Population amalgamation and genetic variation: observations on artificially agglomerated tribal populations of Central and South America. *Am J Hum Genet* 43:709–725
- Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* 144:2001–2014
- Crawford M (1998) The origin of Native Americans: evidence from archeological genetics. Cambridge University Press, Cambridge
- Dillehay T (1999) The late Pleistocene cultures of South America. *Evol Anthropol* 7:206–216
- Dipierri JE, Alfaro E, Martinez-Marignac VL, Bailliet G, Bravi CM, Cejas S, Bianchi NO (1998) Paternal directional mating in two Amerindian subpopulations located at different altitudes in northwestern Argentina. *Hum Biol* 70:1001–1010
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–491
- Forster P, Harding R, Torroni A, Bandelt HJ (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet* 59:935–945
- Forster P, Röhl A, Lünemann P, Brinkmann C, Zerjal T, Tyler-Smith C, Brinkmann B (2000) A short tandem repeat-based phylogeny for the human Y chromosome. *Am J Hum Genet* 67:182–196
- Harlan J (1971) Agricultural origins: centers and non centers. *Science* 174:468–474
- James F, Culloch CM (1990) Multivariate analysis in ecology and systematics: panacea or pandora box? *Ann Rev Ecol Syst* 21:129–166
- Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P, Wiebe V, Griffiths RC, Templeton AR, Hammer MF (1999) Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. *Am J Hum Genet* 64:817–831
- Kayser M, Caglia A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, et al (1997) Evaluation of Y-chromosome STRs: a multicenter study. *Int J Legal Med* 110:125–133
- Kolman CJ, Bermingham E (1997) Mitochondrial and nuclear DNA diversity in the Choco and Chibcha Amerinds of Panama. *Genetics* 147:1289–1302
- Kruskal J (1964) Nonmetric multidimensional scaling: a numerical method. *Psychometrika* 29:28–42
- Luiselli D, Simoni L, Tarazona-Santos E, Pastor S, Pettener D (2000) Genetic structure of Quechua-speakers of Central Andes and geographic patterns of gene frequencies in South Amerindian populations. *Am J Phys Anthropol* 113:5–17
- Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res* 27:209–220
- Maruyama T, Fuerst PA (1985) Population bottlenecks and nonequilibrium models in population genetics. III. Genic homozygosity in populations which experience periodic bottlenecks. *Genetics* 111:691–703
- Merriwether DA, Hall WW, Vahlne A, Ferrell RE (1996) mtDNA variation indicates Mongolia may have been the source for the founding population for the New World. *Am J Hum Genet* 59:204–212
- Mesa NR, Mondragón MC, Soto ID, Parra MV, Duque C, Ortíz-Barrientos D, García LF, Velez ID, Bravo ML, Múnera JG, Bedoya G, Bortolini MC, Ruiz-Linares A (2000) Autosomal, mtDNA, and Y-chromosome diversity in Amerinds: pre- and post- Columbian patterns of gene flow in south America. *Am J Hum Genet* 67:1277–1286
- Neel JV (1978) The population structure of an Amerindian tribe, the Yanomama. *Annu Rev Genet* 12:365–418.

- Nei M (1987) *Molecular evolutionary genetics*. Columbia University Press, New York
- Pena SD, Santos FR, Bianchi NO, Bravi CM, Carnese FR, Rothhammer F, Gerelsaikhon T, Munkhtuja B, Oyunsuren T (1995) A major founder Y-chromosome haplotype in Amerindians. *Nat Genet* 11:15–16
- Pettener D, Pastor S, Tarazona-Santos E (1998) Surnames and genetic structure of a high-altitude Quechua community from the Ichu River Valley, Peruvian Central Andes, 1825–1914. *Hum Biol* 70:865–887
- Powell JF, Neves WA (1999) Craniofacial morphology of the first Americans: pattern and process in the peopling of the New World. *Am J Phys Anthropol* 110 Suppl 29:153–188
- Rickards O, Martínez-Labarga C, Lum JK, De Stefano GF, Cann RL (1999) mtDNA history of the Cayapa Amerinds of Ecuador: detection of additional founding lineages for the Native American populations. *Am J Hum Genet* 65:519–530
- Rickards O, Tartaglia M, Martínez-Labarga C, De Stefano GF (1994) Genetic characterization of the Cayapa Indians of Ecuador and their genetic relationships to other Native American populations. *Hum Biol* 66:299–322
- Ruiz-Linares A, Ortiz-Barrientos D, Figueroa M, Mesa N, Munera JG, Bedoya G, Velez ID, Garcia LF, Perez-Lezaun A, Bertranpetit J, Feldman MW, Goldstein DB (1999) Microsatellites provide evidence for Y chromosome diversity among the founders of the New World. *Proc Natl Acad Sci USA* 96:6312–6317
- Salzano F, Callegari-Jacques S (1988) *South American Indians: a case study in evolution*. Clarendon Press, Oxford
- Sánchez-Albornoz N (1992) La población de la América Colonial Española. In: Bethell L (ed) *Historia de América Latina*. 1. América Latina Colonial: La América Precolombina y la Conquista. Editorial Crítica, Barcelona, pp 15–38
- Sandweiss DH, McInnis H, Burger RL, Cano A, Ojeda B, Paredes R, Sandweiss MC, Glascock MD (1998) Quebrada Jaguay: early South American maritime adaptations. *Science* 281:1830–1832
- Santos FR, Carvalho-Silva DR, Pena SDJ (1999a) PCR-based DNA profiling of human Y chromosomes. In: Epplen JT, Lubjuhn T (eds) *DNA profiling and DNA fingerprinting: methods and tools in biosciences and medicine*. Birkhauser Verlag, Basel, pp 133–147
- Santos FR, Pandya A, Tyler-Smith C, Pena SDJ, Schanfield M, Leonard WR, Osipova L, Crawford MH, Mitchell RJ (1999b) The central Siberian origin for native American Y chromosomes. *Am J Hum Genet* 64:619–628
- Santos FR, Rodriguez-Delfin L, Pena SD, Moore J, Weiss KM (1996) North and South Amerindians may have the same major founder Y chromosome haplotype. *Am J Hum Genet* 58:1369–1370
- Santos FR, Tyler-Smith C (1996) Reading the human Y chromosome: the emerging DNA markers and human genetic history. *Braz J Genet* 19:665–670
- Schneider S, Roessli D, Excoffier L (2000) *Arlequin*, v 2.000: a software for population genetics data analysis. Genetics and Biometry Laboratory. University of Geneva, Geneva
- Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, Weiss KM, Wallace DC (1990) Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. *Am J Hum Genet* 46:613–623
- Simoni L, Tarazona-Santos E, Luiselli D, Pettener D (2000) Genetic differentiation of South American native populations inferred from classical markers: from explorative analysis to a working hypothesis. In: Renfrew C (ed) *America past, America present: genes and languages in the Americas and beyond*. McDonald Institute for Archeological Research, Cambridge, pp 123–134
- Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139:457–462
- Smouse P (1998) To tree or not to tree? *Mol Ecol* 7:399–412
- Smouse P, Long J (1992) Matrix correlation analysis in anthropology and genetics. *Yearb Phys Anthropol* 35:187–213
- Sokal R (1979) Ecological parameters inferred from spatial correlograms. In: Patil G, Rozenzweig M (eds) *Contemporary quantitative ecology and related econometrics*. International Co-operative Publishing, Fairland, MD
- Templeton AR (1998) Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Mol Ecol* 7:381–397
- Torroni A, Sukernik RI, Schurr TG, Starikorskaya YB, Cabell MF, Crawford MH, Comuzzie AG, Wallace DC (1993) mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. *Am J Hum Genet* 53:591–608
- Underhill PA, Jin L, Zemans R, Oefner PJ, Cavalli-Sforza LL (1996) A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. *Proc Natl Acad Sci USA* 93:196–200
- Ward RH, Frazier BL, Dew-Jager K, Pääbo S (1991) Extensive mitochondrial diversity within a single Amerindian tribe. *Proc Natl Acad Sci USA* 88:8720–8724
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370
- Wright S (1943) Isolation by distance. *Genetics* 28:114–138